Two Level Security for Cloud Storage with Data Deduplication

1Pallavi N Marathe, 2 Shivani Deosthale, 3Rashmi Chavan
1,2,3 Information Technology.
Shah and Anchor Kutchhi Engineeing College,
Chembur, Mumbai, India.
sakec.pallavim@gmail.com

Abstract-- Cloud computing provides number of services to client over internet. Storage service is one of the important services that people used now days for storing data on network so that they can access their data from anywhere and anytime. With the benefit of storage service there is an issue of security. To overcome security problem the proposed system contain two levels of securities and to reduce the unwanted storage space de-duplication technique is involved. To increase the level of security one technique is a session password. Session passwords can be used only once and every time a new password is generated. To protect the confidentiality of sensitive data while supporting de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. Symmetric key algorithm uses same key for both encryption and decryption. In this paper, I will focus on session based authentication for login, encryption for files and duplication check for reduce space of storage on cloud.

Keywords- Encryption, data deduplication, authentication.

*****

I. INTRODUCTION

Cloud Storage is a system whereby data is remotely stored, maintained, managed, and backed up. The service is available to users over a network, which is usually the internet. It allows the user to store files online so that the user can access them from any location through the internet. The service providing company makes them available to the user online by keeping the uploaded files on an external server. This gives companies using cloud storage services ease and convenience, but can potentially be costly. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by users with specified privileges. Two critical challenges of cloud storage services are the security and management of the ever-increasing volume of data. Security is basically divide in two levels at the initial point that is session based password for login and for storing files encryption is used to protect data form outsider. The most common method used for authentication is text password. The vulnerabilities of this method like eves dropping, dictionary attack, social engineering and shoulder surfing are well known. Random and lengthy passwords can make the system secure. But the main problem is the difficulty of remembering those passwords. Studies have shown that users tend to pick short passwords or passwords that are easy to remember. Unfortunately, these passwords can be easily guessed or cracked. The alternative techniques are graphical passwords and biometrics. But these two techniques have their own disadvantages. Biometrics, such as finger prints, iris scan or facial recognition have been introduced but not yet widely adopted. The major drawback of this approach is that such systems can be expensive and the identification process can be slow. There are graphical passwords schemes that have been proposed which are resistant to shoulder-surfing. One such method is session password grid for password. The next method is encryption for storing file on cloud. Encryption is a mathematical process of transforming the data with a key. The proposed system uses symmetric key encryption for storing data on file and for increasing more security salt is added in encryption algorithm so key becomes strong. Even though the data is accessed by the third party, they shouldn't get the actual data. So, all the data must be encrypted before it is transmitted to the cloud storage. Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy.

II. RELATED WORKS

P.Anderson and L.Zhang [1] describes an algorithm which takes advantage of the data which is common between users to increase the speed of backups, and reduce the storage requirements. This algorithm supports client-end per-user encryption which is necessary for confidential personal data. It also supports a unique feature which allows immediate detection of common subtrees, avoiding the need to query the backup system for every file. We describe a prototype implementation of this algorithm for Apple OS X, and present an analysis of the potential effectiveness, using real data obtained from a set of typical users. Finally, we discuss the use of this prototype in conjunction with remote cloud storage, and present an analysis of the typical cost savings.

ISSN: 2321-8169
4090 - 4093

M.Bellare, S.Keelveedhi, and T.Ristenpart [2] proposed an architecture that provides secure deduplicated storage resisting brute-force attacks, and realize it in a system called DupLESS. In DupLESS, clients encrypt under message-based keys obtained from a key-server via an oblivious PRF protocol. It enables clients to store encrypted data with an existing service, have the service perform deduplication on their behalf, and yet achieves strong confidentiality guarantees. Encryption for deduplicated storage can achieve performance and space savings close to that of using the storage service with plaintext data.

M. Bellare, S. Keelveedhi, and T. Ristenpart [3] state a new cryptographic primitive, Message-Locked Encryption (MLE), where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure deduplication (space-efficient secure outsourced storage), a goal currently targeted by numerous cloud-storage providers. Author provides definitions both for privacy and for a form of integrity that call tag consistency. On the practical side, ROM security analyses of a natural family of MLE schemes that includes deployed schemes. On the theoretical side the challenge is standard model solutions, and connections with deterministic encryption, hash functions secure on correlated inputs and thesample-then-extract paradigm to deliver schemes under different assumptions and for different classes of message sources.

S.Ruj, M. Stojmenovic and A. Nayak,[4] says In the proposed scheme, the cloud verifies the authenticity of the user without knowing the user's identity before storing information. The scheme also has the added feature of access control in which only valid users are able to decrypt the stored information. The scheme prevents replay attacks and supports creation, modification, and reading data stored in the cloud. Moreover, our authentication and access control scheme is decentralized and robust, unlike other access control schemes designed for clouds which are centralized.

C. Wang, Q. Wang, K. Ren, N. Cao and W. Lou [5] Says in proposed design allows users to audit the cloud storage with very lightweight communication and computation cost. The auditing result not only ensures strong cloud storage correctness guarantee, but also simultaneously achieves fast data error localization, i.e., the identification of misbehaving server. Considering the cloud data are dynamic in nature, the proposed design further supports secure and efficient dynamic operations on outsourced data, including block modification, deletion, and append. Analysis shows the proposed scheme is highly efficient and resilient against Byzantine failure, malicious data modification attack, and even server colluding attacks.

S. Balaji, Lakshmi A., V. Revanth, M. Saragini, V. Venkateswara Reddy[6] In this paper, they proposed two authentication schemes for generating the session passwords which is identified as the primary level of authentication. various comprehensive investigations on the existing authentication schemes have been accomplished. And it has been discerned that none of the recent authentication schemes can resist all sorts of attacks. Literature review reveals all the studies that are done in past. Some of the authentication schemes are discussed as follows:

A. Pair-based authentication scheme

In pair based authentication scheme the length of the password is 8 and it should contain even number of combination of characters and numerics. An interface consists of 6X6 grid. The grid contains both alphabets and numbers, which are placed at random and the interface changes every time. The mechanism is Firstly, the user has to consider the password in terms of pairs. The first letter in the pair is used to select the row and the second letter is used to select the column in the 6X6 grid. The intersection letter of the selected row and column generates the character which is a part of the session password. In this way, the logic is repeated for all other pairs in the password. [6]

Fig1. Pair-based authentication

B. Hybrid textual authentication scheme

In this scheme, there are two grids one for 1 to 8 numbers which are placed in 8X8 grid and second is a color grid is also displayed containing 4 pairs of colors. Both these grids changes for every session. The logic involved in this scheme is that the rating given to the first color of every pair represents a row and the rating given the second color in that pair represents a column of the 8X8 number grid. The number in the intersection of the row and column of the grid is the part of session password. [6]

 ISSN: 2321-8169
4090 - 4093

Fig2. Hybrid textual authentication scheme

III. PROPOSED SYSTEM

The proposed method will focus on deduplication with attribute based encryption and secure session authentication for every user who is login into cloud. So the proposed system provides secure cloud storage for user as well as provide deduplication scheme to reduce amount of storage space.

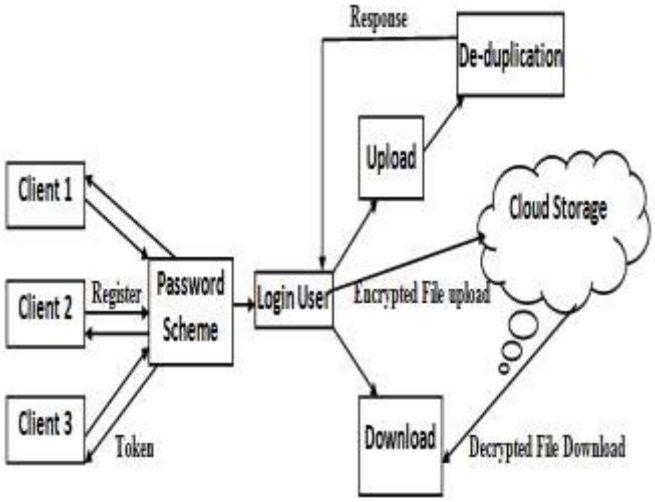

Fig.3 Architecture of proposed system

A. Methodology

Fig1. Shows the architecture of proposed system, In this system first user need to register into the system. For every login one OTP is send to user, for creating password session based authentication scheme is used. After successful login user can upload and download file on cloud. If user wants to upload file deduplication check is perform and then file is uploaded into system if file is not duplicate. For downloading a file, user needs to send request to the owner of the file and once owner accept the request on access key is send to user so user can download the file.

Three main techniques used in system are described below:

1 Deduplication

Data deduplication is one of important data compression techniques for eliminating duplicate copies of repeating data. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication can take place at either the file level or the block level. For file level deduplication, it eliminates duplicate copies of the same file. Deduplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files.

2. Attribute based encryption

Attribute-based encryption (ABE) is a used for to better protect data from unauthorized user. In this encryption technique user attribute like user name, user email-id, file content, extension are used to encrypt and decrypt file on cloud. Symmetric key encryption is used for both file encryption and key generation. For encryption AES algorithm is used and for key generation Rijndael algorithm is used. In this system the receiver receive attribute and secret key if user wants to download a file and able to decrypt file once user got the access key.

3. Session based authentication

Maximum length of the password is 12 and it can be called as secret pass. There is no restriction for password like it contain even as well as odd password. Session passwords are generated based on this secret pass. During the login phase, when the user enters his username, in back end interface create password based on grid. Then the secret pass is send to user on email and phone, this password is called as OTP(One Time Password).The grid is of size 6 x 8 and it consists of alphabets, numbers and special characters. These are randomly placed on the grid and the interface changes every time. User has to enter the password depending upon the secret pass. User has to consider his secret pass in terms of pairs. The session password consists of alphabets, digits and special characters. The first letter in the pair is used to select the both row and column respectively and the second letter is used to select the both column and row. . If password is odd then the first letter of password is append at the end of the password for making pair. The intersection letter is part of the session password. This is repeated for all pairs of secret pass

A	B	!	)	T	X
U	C	D	E	(	^
0	G	1	V	F	&
N	2	#	3	O	M
Y	7	$	Z	4	*
8	Q	9	I	P	_
R	K	@	J	5	H
%	6	L	W	S	+

Fig4. Grid for password authentication

B. Implementation

Implementation includes following steps

- *User Profiling:* New users are supposed to register with their credentials in order to create an account. The user has to enter attributes like name, username, password, email and mobile for registration.

ISSN: 2321-8169
4090 - 4093

- *Session Password Generation:* In order to avoid shoulder surfing attack, new passwords are generated every time a user goes through the login phase. This password as token is mailed to the user through his registered mail address. On successful verification, the user can access his account.
- *Upload File:* To upload a file on cloud user need to rlogin first in to system and then the deduplication check is performed. If file is not duplicate then file is encrypted with Access key and dynamic salt, then it is uploaded on to the cloud.
- *Download file:* If user wants to download the file, request is send to the owner of the file. If file owner accept the request then the secret key is send to user who send request to owner. With that secret key user can download the file.
- *Duplication Check:* This step is to avoid storing files of same content and name. This helps in maintaining the redundancy of the database. The system goes through 3 levels of duplication:
    - ✓ *File Level:* To check the file name and avoid uploading files of same name.
    - ✓ *Byte Level:* To check the file content and avoid uploading files with same content and different names.
    - ✓ *Block Level:* this comes after first two levels, here the file is divided into 5 blocks and then compared with the blocks of already uploaded files. If more than 3 blocks are similar then it is considered as a duplicate file.[7]
- *File Token Generation (Access Key):* file token is generated using the random generator function. This token is used to encrypt and decrypt the file and also used as an access key to download the file.
- *Dynamic Salt:* Salt is nothing but a random value used to generate a new pseudo random key from the file token. This is done to avoid dictionary attack. Every file has a different token and different salt.
- *Encryption:* A new secret key is generated using the token and salt. This key is used to encrypt the attribute of the file. The attribute of the file is nothing but the file content.
- *Decryption:* The secret key used to encrypt the file is also used for the decryption.

IV. CONCLUSION

Authorized data deduplication scheme proposed to protect the data security by including differential privileges of users in the duplicate check. System provide several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. Security analysis demonstrates that schemes are secure in terms of insider and outsider attacks specified in the proposed security model. The authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer. Moreover, the encryption of the file to store in the cloud is done attribute based. Key distribution is done in a decentralized way and also hides the attributes and access policy of a user. For more security, session grid for password authentication is used. This helps to prevent the shoulder surfing attack.

REFERENCES

[1] P. Anderson and L. Zhang, “Fast and secure laptop backups with encrypted de-duplication”, In Proc. of USENIX LISA, 2010.
[2] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Dupless: Server aided encryption for deduplicated storage”, USENIX Security Symposium, 2013.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Message-locked encryption and secure deduplication”, In EUROCRYPT, pages 296– 312, 2013.
[4] S. Ruj, M. Stojmenovic and A. Nayak, “Privacy Preserving Access Control with Authentication for Securing Data in Clouds”, IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pp. 556–563, 2012.
[5] C. Wang, Q. Wang, K. Ren, N. Cao and W. Lou, “Toward Secure and Dependable Storage Services in Cloud Computing”, IEEE T. Services Computing, vol. 5, no. 2, pp. 220–232, 2012.
[6] S. Balaji, Lakshmi A., V. Revanth, M. Saragini, V. Venkateswara Reddy, ”Authentication Techniques for Engendering Session Passwords with Colors and Text”, Advances in Information Technology and Management, Vol. 1, No. 2, 2012.
[7] Nesrine Kaaniche, Maryline Laurent, a secure client side deduplication scheme in cloud storage environments, 6th international conference on new technologies, mobility and security,